

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Automatic Tuning and Software Modularization Techniques for Optimized Large Scale Data Analytics
著者(和文)	MIYURUDAYARATHNA
Author(English)	Miyuru Dayarathna
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第9347号, 授与年月日:2013年9月25日, 学位の種別:課程博士, 審査員:鈴村 豊太郎,徳田 雄洋,横田 治夫,村田 剛志,小林 隆志
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第9347号, Conferred date:2013/9/25, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	Miyuru Dayarathna		
論文審査 審査員		氏 名	職 名		氏 名	職 名
	主査	鈴木豊太郎	客員准教授	審査員	小林隆志	准教授
	審査員	徳田雄洋	教授			
		横田治夫	教授			
		村田剛志	准教授			

論文審査の要旨 (2000 字程度)

本論文は「Automatic Tuning and Software Modularization Techniques for Optimized Large Scale Data Analytics」と題し、主要な3部に及び12章から構成され、英文で記述されている。第1章では、エクサスケールの大規模計算環境に向けて、データストリーム処理および大規模グラフ処理のための新たな性能評価手法、性能最適化手法、ライブラリ、そしてドメイン特化型言語の重要性を述べ、論文の構成を示している。また、第2章では関連研究を述べ、第3章では本研究の基礎となるソフトウェア技術の概要を述べている。

第1部「Stream Processing Optimization」では、データストリーム処理における性能最適化手法の研究を述べている。第4章と第5章では、データストリームの処理量を適応的に調整するフレームワークを提案し、代表的なデータストリーム処理系のシステムアーキテクチャの違いによる性能特性を検証し、次章につながる性能最適化の重要性に関して述べている。第6章、第7章、第8章ではデータストリーム処理アプリケーションの自動性能最適化手法 Hirundo を提案している。Hirundo は、頻出するデータフローパターンと実行時の性能プロファイリングの結果を元にデータフロー変換を適用し、計算資源、アプリケーションの計算特性およびデータ処理特性に応じて最適なデータフローパターンを探索する手法である。この提案手法によって、アプリケーションロジックを開発者が直感的にデータフローとして表現した場合より、実アプリケーションにおいて最大約30倍の性能向上を得られることを検証している。既存の研究では、データフローを構成する複数の処理要素をコンパイルする際に、実行プロファイルを用いて複数の処理要素を適切な粒度のプロセスに束ねる手法 Fusion が提案されているが、提案手法により最適化された性能はこの手法よりも上回ることを示している。

第2部「Graph Processing Optimization」では、エクサスケール環境における大規模グラフ解析のソフトウェアシステムについて述べている。第9章では、大規模かつスケーラブルなグラフデータベースを設計・実装するにあたって指針となるベンチマークシステムを、乗法性属性グラフモデル (MAG モデル) に従って設計している。これによって、クロネッカーグラフと比較し、実世界の属性グラフを適切にモデリングすることが可能なことを示し、複数のグラフデータベースの性能特性を示している。第10章では、10億規模の頂点とエッジを持つ大規模グラフの解析ライブラリ ScaleGraph の設計および実装について述べている。今後のエクサスケールに向けたスーパーコンピュータではメモリーやアクセラレータ、ストレージなどのハードウェアの発展により、それを支えるシステムソフトウェアも変化していく。このような変化に追従するライブラリを設計するため、分散メモリー環境、通信レイヤ、アクセラレータなどを隠蔽することが可能な並列分散言語 X10 を採用し、スケーラブルでかつポータビリティの高いライブラリを提案し、実装した。また、スーパーコンピュータ TSUBAME 2.0 を用いて、1ノードで処理することが困難な大規模グラフを用いて性能検証し、内向辺次数分布、最短経路問題、媒介中心性などのグラフ解析において、ScaleGraph がスケーラブルな性能を実現できることを示している。

次に、第11章では大規模グラフ解析のドメイン特化言語 Exedra を提案している。この言語によって、ユーザーは多様なグラフ処理ライブラリを透過的に利用することが可能になり、ライブラリが採用するグラフデータや解析の内部構造を意識することが不要になる。

第3部「Conclusion and Future Directions」の第12章では、本論文の結論および貢献について述べ、今後の課題についてまとめている。

以上で述べたように、本論文では、データストリーム処理と大規模グラフ処理のスケーラブルなソフトウェアシステムおよび性能最適化手法を提案し、その有効性を実験によって示している。本研究で得られた知見は、大規模データ解析の研究領域において価値が高く、学術上、寄与するところが多い。よって、本論文は博士 (学術) の学位論文として十分価値があるものと認める。